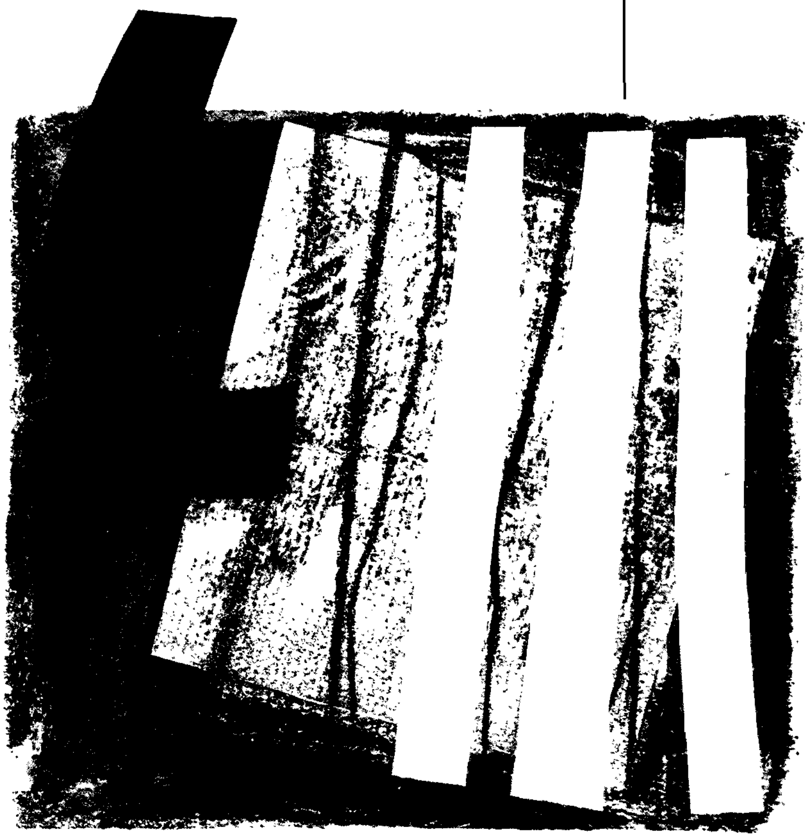
DERIVATIVE ESTIMATION AND TESTING IN GENERALIZED ADDITIVE MODELS

Lijian Yang, Stefan Sperlich and Wolfgang Härdle

00-71



WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 00-71
Statistics and Econometrics Series 34
October 2000

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

DERIVATIVE ESTIMATION AND TESTING IN GENERALIZED ADDITIVE MODELS

Lijian Yang, Stefan Sperlich and Wolfgang Härdle *

Abstract

Estimation and testing procedures for generalized additive (interaction) model are developed. We present extensions of several existing procedures for additive models when the link is the identity. This set of methods includes estimation of all component functions and their derivatives, testing functional forms and in particular variable selection. Theorems and simulation results are presented for the fundamentally new procedures. These comprise of, in particular, the introduction of local polynomial smoothing for this kind of models and the testing, including variable selection. Our method is straightforward to implement and the simulation studies show good performance in even small data sets.

Keywords: Component Analysis; Derivative Estimation; Generalized Additive Models; Local Polynomial Estimator; Marginal Integration.

*Yang, Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA; Sperlich, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, C/ Madrid 126 28903 Getafe, Madrid, Spain, e-mail: stefan@est-econ.uc3m.es; Härdle, Institut für Statistik und Ökonometrie, Wirtschaftswissenschaftliche Fakultät, Humboldt-Universität zu Berlin, D 10178 Berlin, Germany. This research was supported by the Deutsche Forchungsgemeinschaft, SFB 373 "Quantifikation und Simulation ökonomischer Prozesse", Berlin, the Spanish "Dirección General de Enseñanza Superior", number PB98-0025, and NSF grant DMS 9971186. The authors thank Oleg Lepskii for helpful discussion.

1 Introduction

Additive models in non parametric regression analysis are rather popular for mainly two reasons. In economic theory additivity is equivalent to the well known property called strong separability and has many straightforward consequences for analysis, interpretation and decision making. In statistics it is well known due to articles of Stone (1985/86) that additive regression models can be estimated at the univariate rate of convergence. Most flexible model estimators suffer from the so called "curse of dimensionality". This problem disappears if the impact of the regressors $X_1, X_2, \ldots, X_d$ on the response Y is in some sense separable, e.g. when the regression function $E[Y|X] = m(X)$ is additive

$$(1) \qquad m(X) = c + \sum_{\beta=1}^{d} f_\beta(X_\beta).$$

Here, c is a constant and $\{f_\beta(\cdot)\}_{\beta=1}^{d}$ is a set of unknown functions. They are assumed to be smooth but otherwise arbitrary up to the identifiability condition $Ef_\beta(X_\beta) = 0$ for every $1 \leq \beta \leq d$.

Among the existing procedures as backfitting (Hastie and Tibshirani, 1990, Mammen, Linton and Nielsen, 1999) and series estimator (Andrews and Whang, 1990), the marginal integration estimator (Tjøstheim and Auestad, 1994, or Linton and Nielsen, 1995) attracted a fair amount of attention thanks to the appealing simplicity in realization as well as in theory. Further, its interpretation and extension to interactive models is well understood (see Nielsen and Linton, 1997, Sperlich, Linton and Härdle, 1999, or Sperlich, Tjøstheim and Yang, 2000). Although the backfitting is easy to implement, its iterative structure has made its theoretical properties and correctly interpretation difficult, see Mammen, Linton and Nielsen (1999). Moreover, there is no theory for extensions as we know them already for the marginal integration estimator e.g. by Fan, Härdle and Mammen (1998), Severance-Lossin and Sperlich (1999), Linton and Härdle (1996) or Sperlich, Tjøstheim and Yang (2000).

Apparently, model (1) excludes a wide variety of situations. The most natural and often used extension is

$$(2) \qquad G(m(X)) = c + \sum_{\beta=1}^{d} f_\beta(X_\beta),$$

where $G(\cdot)$ is a monotone link function. This is needed for many situations when model (1) is inappropriate, e.g. for binary and survival data. Widely used link functions include the logit and probit links for binary data, and the logarithm transform for Poisson count data. One can also let G be the logarithm function and so the regression function becomes multiplicative. Without loss of generality but along general practice, we assume the link function $G(\cdot)$ to be known a priori. Testing the specification of this link is beyond the scope of this paper but is discussed e.g. in Härdle, Huet, Mammen and Sperlich (2000).

For the generalized additive model (2) (GAM) there is still need for investigation. On the one hand, there is no theory for the many existing backfitting procedures. Derivative estimation is a very important matter, especially in economics, but so far not investigated for these kinds of models. Further, the need of testing methods for various problems as e.g. variable selection, functional forms and additivity is obvious. Since our methods do not restrict the form of link function G, they generalize the work of Hjellvik, Yao and Tjøstheim (1998) which deals exclusively with additive models.

We introduce local polynomial estimation scheme for the components f_β in model (2) and their derivatives. For the ease of notation, asymptotic theory is explicitly derived only for the more complicated case of estimating derivatives. Having these estimates at hand they can be used for testing. This can be either testing against parametric, in our case polynomial, specification or it can be used for variable selection procedures. We construct our test statistics in analog to the one of Härdle and Mammen (1993). We performed a simulation study for the two original new contributions, i.e. derivative estimation and variable selection.

The paper is organized as follows. In the next section, we provide the technical setting for the problems and describes the marginal integration estimators of $f_\alpha^{(\nu)}(\cdot)$. In Section 3 we discuss important extensions. Section 4 presents procedures and theorems for a general testing method. Simulation studies are given in Section 5. All assumptions and proofs are deferred to the appendix.

2 Estimation of Functions and Derivatives

As indicated before the main purpose of this paper is to complete the set of tools for the analysis of marginal impact functions in regression models, especially for generalized additive models with known link function. We will present first procedures and theory for local polynomial smoother in models with possibly non identical link function. For brevity, we give our results in terms of derivative estimation, of which the estimation of component function is a particular case. Before coming closer to the here applied marginal integration method we need some general considerations about derivative estimation in generalized additive regression models.

To make inference on the derivative $f_\alpha^{(\nu)}(\cdot)$, we first want to express it in terms of the known G and the unknown m. Denote the variable $X = (X_\alpha, \overline{X})$ to highlight a particular direction α, where $\overline{X} = (X_1, ..., X_{\alpha-1}, X_{\alpha+1}, ..., X_d)$. The marginal density of X_α, that of $\overline{X}$ and the joint density of $X = (X_\alpha, \overline{X})$ are denoted by $\varphi_\alpha(x_\alpha)$, $\overline{\varphi}(\overline{x})$, and $\varphi(x_\alpha, \overline{x})$ respectively. We define $F_\alpha(x_\alpha) = \int G\{m(x)\}\,\overline{\varphi}(\overline{x})d\overline{x} = c + f_\alpha(x_\alpha)$, for every $1 \le \alpha \le d$, then

$$G\{m(x)\} = \sum_{\alpha=1}^{d} F_\alpha(x_\alpha) - (d-1)c. \tag{3}$$

Taking derivatives on both sides and working by induction on ν gives

Lemma 1 *For $\nu \ge 1$, define $J_\nu = \{(j_1, j_2, \ldots, j_\nu) \mid 0 \le j_1, j_2, \ldots, j_\nu \le \nu$, and $j_1 + 2j_2 + \ldots + \nu j_\nu = \nu\}$, the ν-th derivative $f_\alpha^{(\nu)}(x_\alpha)$ satisfies the following formula*

$$f_\alpha^{(\nu)}(x_\alpha) = \nu! \sum_{(j_1, j_2, \ldots, j_\nu) \in J_\nu} G^{(j_1+j_2+\cdots+j_\nu)}\{m(x_\alpha, \overline{x})\} \prod_{\lambda=1}^{\nu} \frac{\left\{\partial_\alpha^{(\lambda)} m(x_\alpha, \overline{x})\right\}^{j_\lambda}}{(\lambda!)^{j_\lambda} j_\lambda!} \tag{4}$$

where $\partial_\alpha^\lambda m(x) = \frac{\partial^\lambda m(x)}{\partial x_\alpha^\lambda}$.

Note from this lemma that a function of the vector variable x reduces to a function of a scalar variable x_α. Integrating both sides of (4) yields

Lemma 2 *For $\nu \geq 1$*

$$(5) \qquad f_\alpha^{(\nu)}(x_\alpha) = \nu! \sum_{(j_1, j_2, \ldots, j_\nu) \in J_\nu} \int G^{(j_1+j_2+\cdots+j_\nu)}\{m(x_\alpha, \overline{x})\} \prod_{\lambda=1}^{\nu} \frac{\left\{\partial_\alpha^{(\lambda)} m(x_\alpha, \overline{x})\right\}^{j_\lambda}}{(\lambda!)^{j_\lambda} j_\lambda!} \varphi(\overline{x}) d\overline{x}$$

Equation (5) implies that for an i.i.d. sample X_i, $i = 1, 2, \ldots, n$

$$(6) \qquad f_\alpha^{(\nu)}(x_\alpha) = \frac{1}{n} \sum_{i=1}^{n} G^{(j_1+j_2+\cdots+j_\nu)}\left\{m(x_\alpha, \overline{X}_i)\right\} \prod_{\lambda=1}^{\nu} \frac{\left\{\partial_\alpha^{(\lambda)} m(x_\alpha, \overline{X}_i)\right\}^{j_\lambda}}{(\lambda!)^{j_\lambda} j_\lambda!} + O_p(1/\sqrt{n}).$$

This is used in the next paragraph to obtain estimators of $f_\alpha^{(\nu)}(x_\alpha)$ with low dimensional rates typical for the dimension of the considered derivative function. Later we will also introduce a statistic for testing $f_\alpha^{(\nu)}(\cdot) \equiv 0$ based on its estimates.

For statistical inference, let (X_i, Y_i), $i = 1, 2, \ldots, n$ be an i.i.d. sample following model (2). The marginal integration estimator for $f^{(\nu)}(x_\alpha)$, respectively $F_\alpha(x_\alpha)$ from (3) is defined by replacing in equation (5) the unknown expression $m(\cdot)$ by a local polynomial smoother $\tilde{m}(\cdot)$. The integral over the marginal density $\varphi(\bar{x})$ we replace by (marginal) averaging over $\tilde{m}(x_\alpha, \bar{X}_i)$. The multidimensional local polynomial estimator has been introduced in detail by Ruppert and Wand (1994) and by Severance-Lossin and Sperlich (1999) in the context of marginal integration. We need the following notation.

Set for all $l = 1, 2, \ldots, n$ and $\alpha = 0, 1, 2, \ldots, p$, where p is an integer such that $p - \nu > 0$ is odd,

$$Z_\alpha = \left\{(X_{i\alpha} - x_\alpha)^\lambda\right\}_{n \times (p+1)}, \quad W_{l,\alpha} = \mathrm{diag}\left\{\frac{1}{n} K_h(X_{i\alpha} - x_\alpha) L_g(\overline{X}_i - \overline{X}_l)\right\}_{i=1}^{n},$$

$Y = (Y_i)_{n \times 1}$, and E_λ is a $(p+1)$ vector of zeros whose $(\lambda+1)$-element is 1,

K and L are kernel functions, while for any function K, we denote $K_h(u) = K(u/h)/h$, and here h and g are bandwidths. In the following, $K^{(i)}$ denote the i-th convolution of a function K with itself, and $\mu_r(K) = \int u^r K(u) du$. Note that $E_0' \left(Z_\alpha' W_{l,\alpha} Z_\alpha\right)^{-1} Z_\alpha' W_{l,\alpha} Y$ is a special local polynomial smoother to get our $\tilde{m}$.

Now we can give a closed expression for the estimators:

$$\widehat{F}_\alpha(x_\alpha) = n^{-1} \sum_{l=1}^{n} G\left\{E_0' \left(Z_\alpha' W_{l,\alpha} Z_\alpha\right)^{-1} Z_\alpha' W_{l,\alpha} Y\right\},$$

for $F_\alpha(x_\alpha)$ and consequently

$$\widehat{m}(x) = G^{-1}\left\{\sum_{\beta=1}^{d} \widehat{F}_\beta(x_\beta) - (d-1)\frac{1}{nd} \sum_{l=1}^{n} \sum_{\beta=1}^{d} \widehat{F}_\beta(X_{j\beta})\right\},$$

and

$$(7) \qquad \widehat{\partial_\alpha^\lambda m}(x_\alpha, \overline{X}_l) = \lambda! E_\lambda' \left(Z_\alpha' W_{l,\alpha} Z_\alpha\right)^{-1} Z_\alpha' W_{l,\alpha} Y,$$

$$(8) \qquad \widehat{f}_\alpha^{(\nu)}(x_\alpha) = \frac{\nu!}{n} \sum_{l=1}^{n} \sum_{(j_1, j_2, \ldots, j_\nu) \in J_\nu} G^{(\sum_{\lambda=1}^{\nu} j_\lambda)}\left\{\widehat{m}(x_\alpha, \overline{X}_l)\right\} \prod_{\lambda=1}^{\nu} \frac{\left\{\widehat{\partial^{(\lambda)} m}(x_\alpha, \overline{X}_l)\right\}^{j_\lambda}}{(\lambda!)^{j_\lambda} j_\lambda!}.$$

for $m(x)$ and the derivatives $\partial_\alpha^\lambda m(x)$ or $f_\alpha^{(\lambda)}$ respectively. It can be seen easily that in the case of estimating f_α we take

$$\hat{f}_\alpha(x_\alpha) = \widehat{F}_\alpha(x_\alpha) - \frac{1}{nd}\sum_{l=1}^{n}\sum_{\beta=1}^{d}\widehat{F}_\beta(X_{j\beta}).$$

The asymptotics are given in the next theorem.

Theorem 1 *Under assumptions A1 to A6, for any α and for $\nu \geq 1$, the estimated ν-th derivative $\widehat{f}_\alpha^{(\nu)}(x_\alpha)$ satisfies*

$$\sqrt{nh^{2\nu+1}}\left\{\widehat{f}_\alpha^{(\nu)}(x_\alpha) - f_\alpha^{(\nu)}(x_\alpha) - h^{p+1-\nu}b_{\nu\alpha}(x_\alpha)\right\} \xrightarrow{D} N\left\{0, v_{\nu\alpha}(x_\alpha)\right\}$$

where

$$b_{\nu\alpha}(x_\alpha) = \frac{\nu!\mu_{p+1}(K_\nu^*)}{(p+1)!}\int\left\{(G'\circ m)\partial_\alpha^{(p+1)}m\right\}(x_\alpha,\overline{x})\overline{\varphi}(\overline{x}\ )d\overline{x},$$

$$v_{\nu\alpha}(x_\alpha) = (v!)^2\left\|K_\nu^*\right\|_2^2\int\left\{\frac{(G'\circ m)^2\sigma^2}{\varphi}\right\}(x_\alpha,\overline{x})\overline{\varphi}^2(\overline{x})d\overline{x},$$

where $K_\lambda^*(u) = \sum_{t=0}^{p} s_{\lambda t}u^tK(u)$ with $(s_{st})_{0\leq s,t\leq p} = S^{-1} = \{\mu_{s+t}(K)\}_{0\leq s,t\leq p}^{-1}$.

Note here by the definition of matrix S that the λ-th equivalent kernel $K_\lambda^*(u)$ has the following property

$$\mu_q(K_\lambda^*) = \begin{cases} 0 & q\leq p, \quad q\neq\lambda \\ 1 & q=\lambda \\ \Lambda_\lambda & q=p+1 \end{cases}\quad. \tag{9}$$

Now we write

$$\frac{1}{\nu!}\left(\hat{f}_\alpha^\nu(x_\alpha) - f_\alpha^\nu(x_\alpha)\right) = h^{p+1-\nu}\frac{1}{\nu!}b_{\nu\alpha}(x_\alpha) + \sum_{j=1}^{n}w_{j\alpha}\epsilon_j + O_p(\frac{1}{\sqrt{n}} + h^{p+2-\nu}) \tag{10}$$

where

$$w_{j\alpha} = \frac{1}{h^\nu n}K_{\nu h}^*(x_\alpha - X_{j\alpha})\frac{\overline{\varphi}(\overline{X}_j)\sigma(X_j)(G'\circ m)(x_\alpha,\overline{X}_j)}{\varphi(x_\alpha,\overline{X}_j)}.$$

It is easy to verify that this holds whether $h=h_e$ or $h=h_t$, and it will be made use of it in the next sections. We use residuals $Y_i - \widehat{m}(X_i)$, to approximate $\sigma(X_i)\epsilon_i$.

3 Discussion of Extensions

We now discuss briefly possible extensions which allow to consider more general models as done in Section 2. This ordering has been chosen as otherwise the notation would have become much too confusing in Section 2 and especially the Appendix.

So far we have considered the generalized additive model

$$G\{m(x)\} = c + \sum_{\beta=1}^{d} f_\beta(x_\beta) \quad \text{with } E[f_\alpha(X_\alpha)] = 0,$$

where all component functions were univariate. Clearly, the marginal integration idea to estimate marginal impact functions and its derivatives works through also for any other dimension of regressor X_α.

Consequently, the regressors $X_1, \ldots, X_d$ could be grouped into $q \le d$ (non overlapping) groups $Z_1 \in I\!R^{d_1}, \ldots, Z_q \in I\!R^{d_1}$, $\sum_{l=1}^q d_l = d$ ending up in model

$$G\{m(x)\} = c + \sum_{l=1}^{q} g_l(z_l) \quad \text{with } E[g_l(Z_l)] = 0. \tag{11}$$

The common problem is now to find the correct groups. But this question is equivalent to find the significant interactions between the original, univariate regressors $X_1, \ldots, X_d$. For this reason we decompose the regression as follows:

$$G\{m(x)\} = c + \sum_{\alpha=1}^{d} f_\alpha(x_\alpha) + \sum_{\alpha<\beta} f_{\alpha,\beta}(x_\alpha, x_\beta) + \sum_{\alpha<\beta<\gamma}^{d} f_{\alpha,\beta,\gamma}(x_\alpha, x_\beta, x_\gamma) + \ldots.$$

In practice one would stop after the second order interaction to get an idea about the (correct) grouping in equation (11). Therefore the interesting model is usually

$$G\{m(x)\} = c + \sum_{\alpha=1}^{d} f_\alpha(x_\alpha) + \sum_{\alpha<\beta} f_{\alpha,\beta}(x_\alpha, x_\beta) \tag{12}$$

which can be identified when imposing the centering conditions

$$\int f_\alpha(u)\varphi_\alpha(u)du = 0 \qquad \text{and}$$
$$\int f_{\alpha,\beta}(u,v)\varphi_\alpha(u)du = \int f_{\alpha,\beta}(u,v)\varphi_\beta(v)dv = 0.$$

An intensive discussion of the estimation of additive interaction models when the link $G(\cdot)$ is the identity can be found in Sperlich, Tjøstheim and Yang (2000). Therefore we only sketch here the procedure for the case when G is not trivial. Our consideration has been motivated by finding the right grouping in (11), so it is enough to estimate consistently the $f_{\alpha,\beta}$ up to a constant. With the methods presented in Section 4 these estimates can be used for testing significance.

Analogous to F_α we can define $F_{\alpha,\beta}(x_\alpha, x_\beta) = \int G\{m(x_\alpha, x_\beta, \breve{x})\breve{\varphi}(\breve{x})d\breve{x}\}$ where $\breve{x}$ is now the subvector of x containing all elements except x_α, x_β and $\breve{\varphi}$ the marginal density of $\breve{x}$. Some small calculations show that $(F_{\alpha,\beta} - F_\alpha - F_\beta)(\cdot)$ is equal to $f_{\alpha,\beta}$ up to an additive constant. Now we estimate

$$\widehat{F}_{\alpha,\beta}(x_\alpha, x_\beta) = n^{-1} \sum_{l=1}^{n} G\{\widetilde{m}(x_\alpha, x_\beta, \breve{x})\},$$

where $\widetilde{m}$ can be defined similar to above, see Sperlich, Tjøstheim and Yang (2000), and proceed with $(\widehat{F}_{\alpha,\beta} - \widehat{F}_\alpha - \widehat{F}_\beta)(\cdot)$.

4 Hypothesis Testing on Derivatives

We now turn to componentwise testing. The presented procedures will be useful to check significance or such polynomial structure as linearity for the considered functions. The interest in testing whether a function is significant at all is obvious as it enables us to perform variable selection as well as looking for interaction, see Section 3. Testing polynomial structure is motivated by both economic and statistic arguments. Especially linearity has many important consequences in economics and thus is an important assumption to check. On the other hand, if a wanted parametric specification can not be rejected, the empirical researcher will always prefer to use it. This is due to interpretation, facilities in modeling, etc.

As in the preceding sections we will condense the presentation on the case of test statistics with one dimensional derivative functions. Let us first specify the hypothesis we focus on. We want to test the null hypothesis

$$H_0 : \int f_\alpha^{(\nu)}(x_\alpha)^2 \pi(x_\alpha) dx_\alpha = 0 \text{ vs. local alternatives } H_n : \frac{1}{\nu!^2} \int f_\alpha^{(\nu)}(x_\alpha)^2 \pi(x_\alpha) dx_\alpha > C\rho_n,$$

where $\pi(x)$ is a weight function with Lipschitz continuous $(p+1)$-th derivative, $\rho_n = n^{-1} h^{-(2\nu+1/2)}$ and C is

$$(z_{1-\alpha_I} + z_{1-\alpha_{II}})\sigma_T \tag{13}$$

where

$$\sigma_T^2 = \frac{\left\| K_\nu^{*(2)} \right\|_{L^2}^2}{2} \int \left[\int \left\{ \frac{(G' \circ m)^2 \sigma^2}{\varphi} \right\} (x_\alpha, \overline{x}) \overline{\varphi}^2(\overline{x}) d\overline{x} \right]^2 \pi(x_\alpha)^2 dx_\alpha. \tag{14}$$

Further, $z_{1-\alpha_I}$ is the upper $(1-\alpha_I)$-th point of the standard normal variable, $\alpha_I \in (0,1)$ is the prespecified significance level, while α_{II} is the prespecified type II error. We define the test statistic

$$T = \int \frac{1}{(\nu!)^2} \widehat{f}_\alpha^{(\nu)}(x_\alpha)^2 \pi(x_\alpha) dx_\alpha, \tag{15}$$

which is an estimate for $\frac{1}{(\nu!)^2} \int f_\alpha^{(\nu)}(x_\alpha)^2 \pi(x_\alpha) dx_\alpha$. The next theorem will show that T is a suitable statistic for testing H_0.

Theorem 2 *For any given α and $h = h_t = h_0 n^{-\frac{2}{p+3\nu+2}}$ as specified in A2, under assumptions A1-A6, the limiting distribution of T is*

$$\begin{aligned} & h^{2\nu+1/2} nT - h^{2\nu+1/2} n \frac{1}{(\nu!)^2} \int f_\alpha^{(\nu)}(x_\alpha)^2 \pi(x_\alpha) dx_\alpha - \\ & h^{-1/2} K_\nu^{*(2)}(0) \int \left\{ \frac{(G' \circ m)^2 \sigma^2}{\varphi} \right\} (x_\alpha, \overline{x}) \overline{\varphi}(\overline{x})^2 \pi(x_\alpha) dx_\alpha d\overline{x} \\ & - \frac{n h^{p+\nu+3/2} \mu_{p+1}(K^*)}{\nu!(p+1)!} \int \left\{ (G' \circ m) \partial_\alpha^{(p+1)} m \right\} (x_\alpha, \overline{x}) \overline{\varphi}(\overline{x}) f_\alpha^{(\nu)}(x_\alpha) \pi(x_\alpha) dx_\alpha d\overline{x} \\ & \xrightarrow{D} N(0, \sigma_T^2). \end{aligned} \tag{16}$$

The test rule is to reject H_0 if

$$h^{2\nu+1/2} nT \geq h^{-1/2} K_\nu^{*(2)}(0) \int \left\{ \frac{(G' \circ m)^2 \sigma^2}{\varphi} \right\} (x_\alpha, \overline{x}) \overline{\varphi}(\overline{x})^2 \pi(x_\alpha) dx_\alpha d\overline{x} + z_{1-\alpha_I} \sigma_T. \tag{17}$$

The probability of type II error is smaller than α_{II} as $n \to \infty$: for any function $f_\alpha(x)$,

$$P\left[H_0 \text{ is retained } \mid H_n \text{ is true}\right] \leq \alpha_{II} + o(n^{-\frac{p-\nu+1}{p+3\nu+2}})$$

where the term $o(n^{-\frac{p-\nu+1}{p+3\nu+2}})$ implicitly depends on $f_\alpha()$.

Consider the test problem $H_0: f_\alpha$ *is linear*. Note that if looking at $H_0: \int f_\alpha^{(2)}(x_\alpha)^2\pi(x_\alpha)dx_\alpha = 0$, taking $p=3$, then $n^{-\frac{p-\nu+1}{p+3\nu+2}} = n^{-\frac{2}{11}}$ is the rate. Alternatively, we could look on the first derivative, i.e. $H_0: \int f_\alpha^{(2)}(x_\alpha)^2\pi(x_\alpha)dx_\alpha = const.$ Then, with $\nu=1$, $p=2$ we get even a rate of $n^{-\frac{2}{7}}$ although testing against zero or against a constant is basically the same. Apart from the rate, in small samples it can often be preferable for numerical reasons to look on lower degrees of derivatives if possible.

5 Some Simulation Results

We investigated the performance of our procedures in finite samples; first for the derivative and function estimation, then for the variable selection, i.e. component wise testing for significance of the impact functions.

Function and Derivative Estimation

Although the introduction of a nontrivial link function $G(\cdot)$ looks straight forward for the marginal integration, in practice it unfortunately can cause strong negative effects on the small sample performance. We will illustrate this in the following by doing the same simulations twice, one for the identical link and one for $G(\cdot)=\ln(\cdot)$.

We drew $n=200$ independent variables $X \sim U[-2,2]^3$ and considered the models

$$G_\gamma\{m(X)\} = c + \sum_{\alpha=1}^{3} f_\alpha(X_\alpha),\ \gamma=1,2 \quad \text{with} \quad f_1(X_1) = 1.5\sin(-1.5X_1)$$

$$(18) \qquad f_2(X_2) = X_2^2 - \frac{4}{3} \quad , \quad f_3(X_3) = X_3$$

and $c=3$. Further, G_1 is the identity and G_2 the logarithm. Finally we added a standard normal disturbance ε to equation (18).

To get $\tilde{m}$ we used local linear smoother. An intensive discussion about bandwidth choice in the context of marginal integration is given in Severance-Lossin and Sperlich (1999) and Sperlich, Linton and Härdle (1999). However, the problem is that the optimal bandwidth differs not only for each direction but also for the different G_γ and for the different problems of function, respectively derivative estimation. To simplify the investigation we took only one bandwidth vector for the estimation with G_1: $h_1 = (1.0, 1.5, 2.0)$ and one with G_2: $h_2 = (1.5, 1.75, 2.0)$. These are compromises between the wanted smoothness for the different curvatures. Certainly, using a local linear smoother, f_3 could be estimated optimally by a strong oversmoothing in that particular direction. But this would presume prior knowledge of the function form. For the nuisance directions we set always $g=h$.

After running 500 repetitions we had to skip about 1% of the results which suffered from numerical problems when the link was G_2. In Figure 1 (for G_1) and 2 (for G_2) are given the data generating functions, respectively their derivatives, as dotted lines together with the 99% confidence bands for the estimator resulting from the 500 repetitions. Note that we did no bias reduction here. For that reason the also plotted real functions (solid line) can not necessarily expected inside the bands. Moreover, we will see the structural biases appearing without undersmoothing or other bias reduction methods.

Figure 1 about here

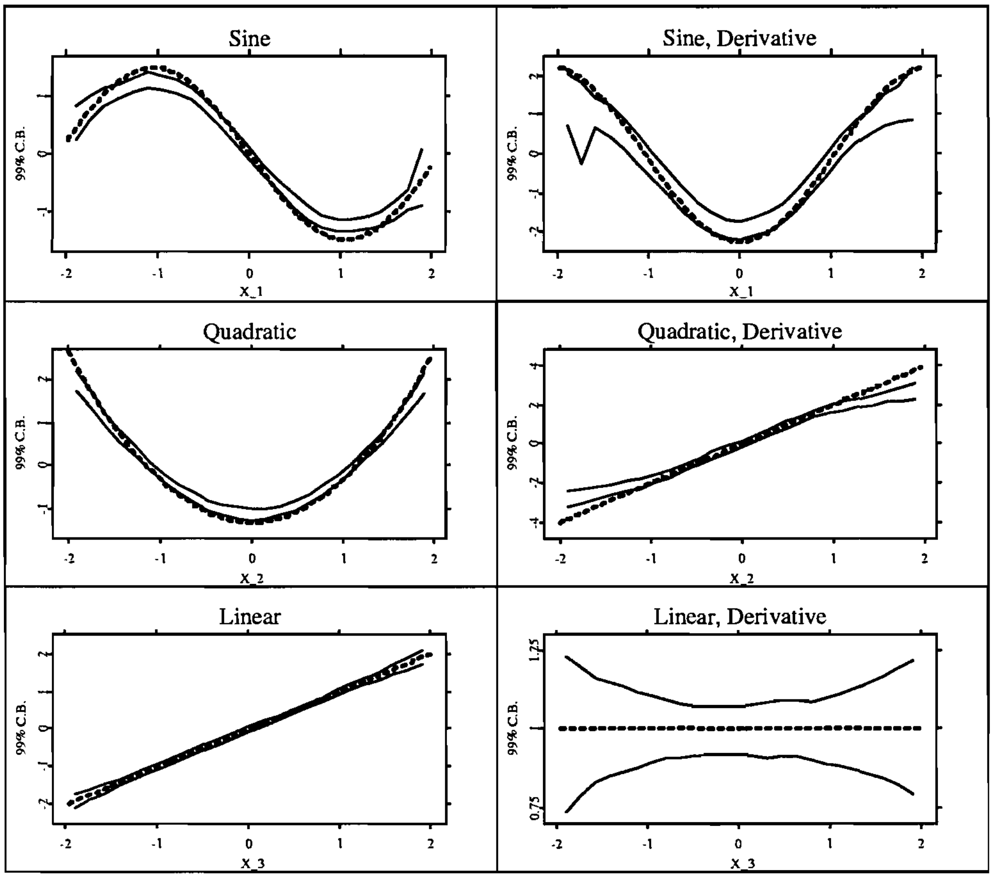


Figure 1: Model (18) with identity link G_1. Functions on the left, derivatives on the right. Dotted lines are the data generating functions, respectively their derivatives, solid lines are the 99% confidence bands after 500 runs.

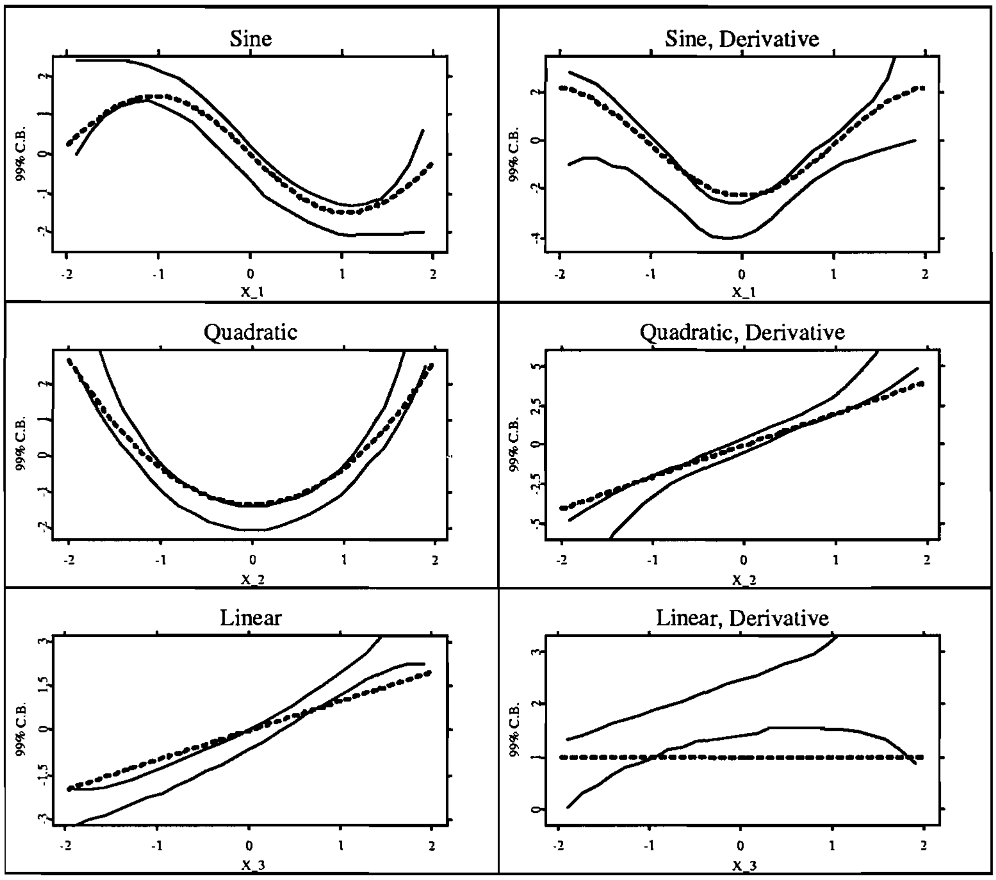


Figure 2: Model (18) with log link G_2. Functions on the left, derivatives on the right. Dotted lines are the data generating functions, respectively their derivatives, solid lines are the 99% confidence bands after 500 runs.

Figure 2 about here

Though the procedures seem to work reasonably well, we recognize an enormous loss of exactness when the link is not trivial. Not surprisingly, the derivative estimation with only $n = 200$ observations seems to be pretty hard, especially for G_2. We can recognize further the biases and boundary effects. As indicated before, the chosen bandwidthes do not seem to be optimal but reasonable.

Testing the Component Functions

It is well known that the asymptotics derived in Theorem 2 are not very helpful when applying the test in practice on small samples. Instead, usually wild bootstrap (see e.g. Liu, 1988 or Wu, 1986) is used to better approximate the distribution of the statistic T under hypothesis H_0. A detailed discussion of this in the context of marginal integration can be found in Sperlich, Tjøstheim and Yang (2000) or Gozalo and Linton (1999).

As the testing problem is easier than estimation, we consider here a more complicated model:

$$G_T\{m(X)\} = \sum_{\alpha=1}^{3} f_\alpha(X_\alpha) \quad \text{with } f_3(X_3) = a \cdot X_3 \tag{19}$$

with f_1, f_2 as in (18) and $G_T(u) = -\ln(\frac{1}{u} - 1)$. So we observe

$$Y = \begin{cases} 1 & \text{if } \sum_{\alpha=1}^{3} f_\alpha(X_\alpha) > \varepsilon \\ 0 & \text{else} \end{cases},$$

where ε logit distributed. Again we drew $n = 200$ independent variables $X \sim U[-2,2]^3$

To implement the test, T is computed as

$$\hat{T} = \frac{1}{n}\sum_{j=1}^{n} \frac{\widehat{f_\alpha^{(\nu)}}(X_{j\alpha})^2 \pi(X_{j\alpha})}{\frac{1}{n}\sum_{t=1}^{n} K_h(X_{j\alpha} - X_{t\alpha})} \frac{1}{(\nu!)^2} \tag{20}$$

For the wild bootstrap, take observations Y_i^*, $i = 1, \ldots, n$ drawn from the (estimated) data generating process under H_0, given $(X_i)_{i=1}^n$. Then calculate the (bootstrap) test statistic $\hat{T}^*$ out from sample $(X_i, Y_i^*)_{i=1}^n$. For this (pre-) estimation undersmoothing is recommended, see Härdle and Marron (1991). In our simulation study we used local linear smoother with $h = g = 1.5$ for all directions to estimate the data generating process under H_0. We drew only 249 bootstrap samples to approximate the distribution of $\hat{T}$.

Our aim is to test $H_0 : f_3 \equiv 0$ for increasing a, see (19). We first compare the test statistics based on function estimates with the one based on derivative estimates. It is certainly known that the one based on derivatives is especially of interest when the considered function is not smooth, e.g. has a peek or a jump. On the other hand it is also known that in those cases kernel smoother can not be recommended anyway.

In Theorem 2, for a first derivative based test the local quadratic smoother has been suggested. For those, larger bandwidths are necessary. Therefore we took local quadratic smoother with bandwidth $h = g = 3.0$. In Table 1 the relative frequencies of rejections for function based ($\nu = 0$) as well as for derivative based ($\nu = 1$) tests are given, all after 500 repetitions. Additionally, in Table 2 we give the corresponding variances over the 500 repetitions.

significance level:		1%	5%	10%	15%	p-value
$f_3(u) = 0$	$\nu = 0$	6.0	10.0	14.8	20.8	48.7
	$\nu = 1$	2.0	2.8	6.4	8.4	55.0
$f_3(u) = u$	$\nu = 0$	100	100	100	100	0.0
	$\nu = 1$	24	42	49	58	16.5

Table 1: Relative rejection frequencies and p-values for testing $H_0 : f_3 \equiv 0$ with tests based on function estimate ($\nu = 0$) and based on derivative estimates ($\nu = 1$), using local quadratic smoother with $h = g = 3.0$.

significance level:		1%	5%	10%	15%	p-value
$f_3(u) = 0$	$\nu = 0$	5.7	9.0	12.7	16.5	9.6
	$\nu = 1$	2.0	2.7	6.0	7.7	7.6
$f_3(u) = u$	$\nu = 0$	0.0	0.0	0.0	0.0	0.0
	$\nu = 1$	17.9	24.5	25.1	24.5	3.4

Table 2: Variances for the relative rejection frequencies and p-values for testing $H_0 : f_3 \equiv 0$ with tests based on function estimate ($\nu = 0$) and based on derivative estimates ($\nu = 1$), see Table 1 .

Table 1 about here

Table 2 about here

We tried thereby other bandwidths and different degrees for the local polynomial smoother. We found that for $n = 200$ the bandwidth choice can be very crucial when using local quadratic or higher order polynomials. It can also be seen in Table 1 that though the p-value is fitted well under H_0, the quantiles are not. Thus we conclude that our estimates are just too wiggly or, the data are too sparse causing numerical problems. A very intensive simulation study would be necessary to investigate in detail the performance and usefulness of derivative based tests in this context if the sample size is "small". We see e.g. in Table 1 that it is pretty conservative for these samples.

There are much more encouraging findings for tests based on the function estimate ($\nu = 0$). We present for this case also results when using local linear smoother with bandwidth $h = g = 1.75$. Again we did a simulation study of 500 repetitions. In Table 3 it can be seen how fast the power increases with a from (19).

Table 3 about here

We conclude all in all that even for small data sets but pretty complex model structures our procedures work reasonable well. The function and derivative estimates give clearly the wanted functional forms. The test procedures are more crucial. For such small samples we recommend to use only statistics based on function estimate (if possible) and low polynomial degrees. They perform well in both, fitting the correct quantiles under the hypothesis and showing strong power against the alternative.

		significance level in %				
		1	5	10	15	p-value
	0.00	1.2	4.4	8.4	11.8	54.4
	0.25	6.4	16.4	28.0	36.8	34.5
a	0.50	37.2	60.0	71.6	78.4	10.4
	0.75	84.0	93.6	97.2	98.8	0.01
	1.00	98.4	99.6	100	100	0.00

Table 3: Relative rejection frequencies and p-value for testing $H_0 : f_3 \equiv 0$ with tests based on function estimate, using local linear smoother with $h = g = 1.75$. Left column refers to the alternative $f_3(u) = a \cdot u$.

Appendix

The following assumptions are used

A1: *The kernel $K(\cdot)$ is symmetric, compactly supported and Lipschitz continuous probability density; while the kernel $L(\cdot)$ is symmetric, compactly supported and Lipschitz continuous with $\int L(u)\,du = 1$ and order q* where $q > \frac{1}{4}(d-1)$ *for estimation and* $q > (d-1)\frac{p+1}{p+3\nu}$ *for testing hypotheses (which in effects, can even be relaxed to* $q > (d-1)\frac{p+1-\nu}{p+3\nu}$ *as one can see from the proof);*

A2: *Bandwidths satisfy* $\frac{nhg^{d-1}}{\ln(n)} \longrightarrow \infty$, $\frac{g^{2q}}{h^{p+1}} \to 0$ *and* $h = h_e = h_0 n^{\frac{-1}{2p+3}}$ *for estimation in Section 2 and* $h = h_t = h_0 n^{\frac{-2}{p+3\nu+2}}$ *for testing hypotheses in Section 4.*

A3: *The functions $f_s(\cdot)$'s have bounded Lipschitz continuous* $(p+1)$-th *derivatives.*

A4: *The variance function, $\sigma^2(\cdot)$, is bounded and Lipschitz continuous.*

A5: *φ and $\overline{\varphi}$ are uniformly bounded away from zero and infinity and have bounded Lipschitz continuous $(p+1)$-th derivatives.*

A6: *G is uniformly bounded away from zero and infinity and have bounded Lipschitz continuous $(p+1)$-th derivative.*

Our estimation procedure makes uses of the following lemma which is a generalization of the result of Linton and Härdle (1996).

Lemma A.1 *Under assumptions A1-A6, for any α*

$$\sqrt{nh}(\widehat{F}_\alpha(x_\alpha) - F_\alpha(x_\alpha) - h^{p+1}b_\alpha(x_\alpha)) \xrightarrow{D} N\{0, v_\alpha(x_\alpha)\}$$

where

$$\begin{aligned} b_\alpha(x_\alpha) &= \frac{\mu_{p+1}(K_0^*)}{(p+1)!}\int \left[\frac{(G'\circ m)\left\{\partial_\alpha^{(p+1)}(m\varphi) - m\partial_\alpha^{(p+1)}(\varphi)\right\}}{\varphi}\right](x_\alpha,\overline{x})\overline{\varphi}(\overline{x})d\overline{x} \\ v_\alpha(x_\alpha) &= \|K_0^*\|_2^2\int\left\{\frac{(G'\circ m)^2\sigma^2}{\varphi}\right\}(x_\alpha,\overline{x})\overline{\varphi}^2(\overline{x})d\overline{x}. \end{aligned}$$

Furthermore

(A.1)
$$\sqrt{nh}(\widehat{m}(x) - m(x) - h^{p+1}b(x)) \xrightarrow{D} N\{0, v(x)\}$$

where

$$b(x) = (G^{-1})' \circ G \circ m(x) \sum_{\alpha=1}^{d} b_\alpha(x_\alpha)$$

and

$$v(x) = \left\{(G^{-1})' \circ G \circ m(x)\right\}^2 \sum_{\alpha=1}^{d} v_\alpha(x_\alpha).$$

Proof. It follows the same asymptotic reasoning as Linton and Härdle (1996) with minor changes because of the use of equivalent kernel K_0^* instead of K. The bias here is of order h^{p+1} instead of h^2. Q.E.D.

Proof of Theorem 1. The steps are similar to Severance-Lossin and Sperlich (1999). The special features here is the use of the formula (4) and its empirical version (8), and the fact that for $\lambda = 0, 1, 2, ..., \nu$, the partial derivative estimates $(\widehat{\partial^{(\lambda)} m}(x_\alpha, \overline{X}_l)$ have the bias rates of $h^{p+1-\lambda}$ and variance rates of $\frac{1}{nh^{2\lambda+1}}$, which together with the previous lemma, gives that

$$\sqrt{nh^{2\nu+1}}\frac{1}{\nu!}(\widehat{f}_\alpha^{(\nu)}(x_\alpha) - f_\alpha^{(\nu)}(x_\alpha)) =$$

$$n^{-1}\sum_{l=1}^{n} G'(m(x_\alpha, \overline{X}_l)) \left(\widehat{\partial^{(\nu)} m}(x_\alpha, \overline{X}_l) - \partial^{(\nu)} m(x_\alpha, \overline{X}_l)\right)$$

$$+O(\sqrt{nh^{2\nu+1}}h^{p+2-\nu} + h)$$

where the asymptotics of $n^{-1}\sum_{l=1}^{n}(G' o m)\left(\widehat{\partial^{(\nu)} m} - \partial^{(\nu)} m\right)(x_\alpha, \overline{X}_l)$ is treated as in Severance-Lossin and Sperlich (1999). Q.E.D.

The proof of Theorem 2 is essentially the same with trivial link or more general links. Therefore, to simplify notation, we give the proof in the case of trivial link function. In this case, $G' \circ m \equiv 1$ and $m(X) = c + \sum_{\beta=1}^{d} f_\beta(X_\beta)$. Therefore

$$b_{\nu\alpha}(x) = \frac{\nu! \mu_{p+1}(K_\nu^*)}{(p+1)!} \int \left\{\partial_\alpha^{(p+1)} m\right\}(x_\alpha, \overline{x})\overline{\varphi}(\overline{x})d\overline{x} \; = \frac{\nu! \mu_{p+1}(K^*)}{(p+1)!} f_\alpha^{(p+1)}(x_\alpha)$$

and the expression for T is

$$\int \left\{ \frac{f_\alpha^{(\nu)}(x_\alpha)}{\nu!} + \frac{h^{p+1-\nu}\mu_{p+1}(K^*)}{(p+1)!} f_\alpha^{(p+1)}(x_\alpha) + \sum_{j=1}^{n} w_{j\alpha}\epsilon_j + O_p\left(\frac{1}{\sqrt{n}} + h^{p+2-\nu}\right)\right\}^2 \pi(x_\alpha)dx_\alpha$$

$$= \int \left\{ \frac{f_\alpha^{(\nu)}(x_\alpha)}{\nu!} + \frac{h^{p+1-\nu}\mu_{p+1}(K^*)}{(p+1)!} f_\alpha^{(p+1)}(x_\alpha) + \sum_{j=1}^{n} w_{j\alpha}\epsilon_j \right\}^2 \pi(x_\alpha)dx_\alpha + O_p\left(\frac{1}{n} + h^{2p+4-2\nu}\right)$$

which can be reduced to

$$= Q + \int \frac{f_\alpha^{(\nu)}(x_\alpha)^2}{(\nu!)^2}\pi(x_\alpha)dx_\alpha + \frac{h^{p+1-\nu}\mu_{p+1}(K^*)}{(p+1)!\nu!} \int f_\alpha^{(p+1)}(x_\alpha)f_\alpha^{(\nu)}(x_\alpha)\pi(x_\alpha)dx_\alpha + O_p\left(h^{2p+4-2\nu}\right)$$
(A.2)

with the quadratic term $Q = \int \left\{ \sum_{j=1}^{n} w_{j\alpha}\epsilon_j \right\}^2 \pi(x_\alpha) dx_\alpha$. We leave out here the routine verification that the following cross term is negligible

$$\int 2 \left\{ \frac{f_\alpha^{(\nu)}(x_\alpha)}{\nu!} + \frac{h^{p+1-\nu}\mu_{p+1}(K^*)}{(p+1)!} f_\alpha^{(p+1)}(x_\alpha) \right\} \sum_{j=1}^{n} w_{j\alpha}\epsilon_j \pi(x_\alpha) dx_\alpha .$$

The formula of σ_T^2 in the case of trivial link is simplified to

$$\sigma_T^2 = \frac{\left\| K_\nu^{*(2)} \right\|_{L^2}^2}{2} \int \left\{ \int \frac{\sigma^2(x)\overline{\varphi}^2(\overline{x})}{\varphi(x)} d\overline{x} \right\}^2 \pi^2(x_\alpha) dx_\alpha. \tag{A.3}$$

To derive the asymptotics of Q, write it as $\sum_{j,k=1}^{n} \epsilon_j\epsilon_k A(X_j, X_k)\, \sigma(X_j)\sigma(X_k)$ where

$$A(X_j, X_k) = \frac{1}{h^{2\nu}n^2} \int K_{\nu h}^*(x_\alpha - X_{j\alpha}) K_{\nu h}^*(x_\alpha - X_{k\alpha}) \frac{\overline{\varphi}(\overline{X}_j)\overline{\varphi}(\overline{X}_k)}{\varphi(x_\alpha, \overline{X}_j)\varphi(x_\alpha, \overline{X}_k)} \pi(x_\alpha) dx_\alpha. \tag{A.4}$$

Separating the diagonal and the cross terms, one gets $Q = Q_1 + Q_2$ with

$$Q_1 = \sum_{j=1}^{n} \epsilon_j^2 A(X_j, X_j)\sigma(X_j)\sigma(X_j) = \sum_{j=1}^{n} A(X_j, X_j) \left\{ Y_j - m(X_j) \right\}^2 \tag{A.5}$$

and

$$Q_2 = \sum_{1 \le j < k = n} 2\epsilon_j\epsilon_k A(X_j, X_k)\sigma(X_j)\sigma(X_k)$$

$$= \sum_{1 \le j < k = n} 2A(X_j, X_k) \left\{ Y_j - m(X_j) \right\} \left\{ Y_k - m(X_k) \right\}.$$

We simplify the expressions $A(X_j, X_k)$ and Q_1 in the following lemmata.

Lemma A.2 *$A(X_j, X_k)$ from (A.4) can be written as*

$$\frac{1}{h^{2\nu+1}n^2} (K_\nu^* * K_\nu^*) \left(\frac{X_{j\alpha} - X_{k\alpha}}{h} \right) \frac{\overline{\varphi}(\overline{X}_j)\overline{\varphi}(\overline{X}_k)}{\varphi(X_{j\alpha}, \overline{X}_j)\varphi(X_{j\alpha}, \overline{X}_k)} \pi(X_{j\alpha}) \left\{ 1 + O_p(h) \right\} \tag{A.6}$$

Proof. By definition

$$A(X_j, X_k) = \frac{1}{h^{2\nu}n^2} \int K_{\nu h}^*(x_\alpha - X_{j\alpha}) K_{\nu h}^*(x_\alpha - X_{k\alpha}) \frac{\overline{\varphi}(\overline{X}_j)\overline{\varphi}(\overline{X}_k)}{\varphi(x_\alpha, \overline{X}_j)\varphi(x_\alpha, \overline{X}_k)} \pi(x_\alpha) dx_\alpha$$

$$= \frac{1}{h^{2\nu+1}n^2} \int K_\nu^*(u) K_\nu^* \left(u + \frac{X_{j\alpha} - X_{k\alpha}}{h} \right) \frac{\overline{\varphi}(\overline{X}_j)\overline{\varphi}(\overline{X}_k)}{\varphi(X_{j\alpha} + hu, \overline{X}_j)\varphi(X_{j\alpha} + hu, \overline{X}_k)} \pi(X_{j\alpha} + hu) du$$

$$= \frac{1}{h^{2\nu+1}n^2} (K_\nu^* * K_\nu^*) \left(\frac{X_{j\alpha} - X_{k\alpha}}{h} \right) \frac{\overline{\varphi}(\overline{X}_j)\overline{\varphi}(\overline{X}_k)}{\varphi(X_{j\alpha}, \overline{X}_j)\varphi(X_{j\alpha}, \overline{X}_k)} \pi(X_{j\alpha}) \left\{ 1 + O_p(h) \right\}$$

Lemma A.3 *As $n \to \infty$ it holds in (A.5) that*

$$Q_1 = \frac{(K_\nu^* * K_\nu^*)(0)}{h^{2\nu+1}n} \int \frac{\overline{\varphi}(\overline{x})^2 \sigma(x)^2}{\varphi(x)} \pi(x_\alpha) dx_\alpha d\overline{x} + O_p\left(\frac{1}{h^{2\nu}n} + \frac{1}{h^{2\nu+1}n^{3/2}} \right) \tag{A.7}$$

Proof. We calculate the mean and the variance of Q_1

$$\begin{aligned}
EQ_1 &= nE\left\{A(X_1,X_1)\sigma(X_1)^2\right\} \\
&= nE\frac{1}{h^{2\nu+1}n^2}(K_\nu^* * K_\nu^*)\left(\frac{X_{1\alpha}-X_{1\alpha}}{h}\right)\frac{\overline{\varphi}(\overline{X}_1)^2\sigma(X_1)^2}{\varphi(X_1)^2}\pi(X_{1\alpha})\left\{1+O(h)\right\} \\
&= \int\frac{1}{h^{2\nu+1}n}(K_\nu^* * K_\nu^*)(0)\frac{\overline{\varphi}(\overline{x})^2\sigma(x)^2}{\varphi(x)^2}\pi(x_\alpha)\varphi(x)dx_\alpha d\overline{x}\ \left\{1+O(h)\right\} \\
&= \frac{(K_\nu^* * K_\nu^*)(0)}{h^{2\nu+1}n}\int\frac{\overline{\varphi}(\overline{x})^2\sigma(x)^2}{\varphi(x)}\pi(x_\alpha)dx_\alpha d\overline{x}\ \left\{1+O(h)\right\}
\end{aligned}$$

and

$$Var(Q_1) = nVar\left\{A(X_1,X_1)\sigma(X_1)^2\right\} \leq nE\left\{A(X_1,X_1)^2\sigma(X_1)^4\right\} =$$

$$nE\left\{\frac{1}{h^{2\nu+1}n^2}(K_\nu^* * K_\nu^*)\left(\frac{X_{1\alpha}-X_{1\alpha}}{h}\right)\frac{\overline{\varphi}(\overline{X}_1)^2\sigma(X_1)^2}{\varphi(X_1)^2}\pi(X_{1\alpha})\right\}^2\left\{1+O(h)\right\}$$

$$= \frac{(K_\nu^* * K_\nu^*)(0)^2}{h^{4\nu+2}n^3}E\frac{\overline{\varphi}(\overline{X}_1)^4\sigma(X_1)^4}{\varphi(X_1)^4}\pi^2(X_{1\alpha})\left\{1+O(h)\right\} = O_p\left(\frac{1}{h^{2\nu+1}n^{3/2}}\right)$$

Therefore

$$Q_1 = \frac{(K_\nu^* * K_\nu^*)(0)}{h^{2\nu+1}n}\int\frac{\overline{\varphi}(\overline{x})^2\sigma(x)^2}{\varphi(x)}\pi(x_\alpha)dx_\alpha d\overline{x}\ + O_p\left(\frac{1}{h^{2\nu}n}+\frac{1}{h^{2\nu+1}n^{3/2}}\right)$$

as is in (A.7). Q.E.D.

Note because $E[\epsilon_i] = 0$, $i = 1,2,\ldots,n$ and the random vectors (X_i,ϵ_i), $i = 1,2,\ldots,n$ are i.i.d., Q_2 is an U -statistic, symmetric and *non-degenerate* because $E_j\epsilon_j\epsilon_k A(X_j,X_k)\sigma(X_j)\sigma(X_k) = 0$, where $E_j = E_{\epsilon_j,X_j}$. To apply central limit theorem to this U -statistic, we calculate the following three quantities

1. The variance of one term: $A_n = E\left[\epsilon_1\epsilon_2 A(X_1,X_2)\sigma(X_1)\sigma(X_2)\right]^2$
2. The fourth moment of one term: $B_n = E\left[\epsilon_1\epsilon_2 A(X_1,X_2)\sigma(X_1)\sigma(X_2)\right]^4$
3. The $C_n = E\left[J_n(\epsilon_1,X_1,\epsilon_2,X_2)\right]^2$, where

$$J_n(\epsilon,X,\delta,Y) = E_1\left[\epsilon_1\epsilon A(X_1,X)\sigma(X_1)\sigma(X)\epsilon_1\delta A(X_1,Y)\sigma(X_1)\sigma(Y)\right]$$

and then verify that

(A.8) $$\frac{C_n+\frac{1}{n}B_n}{A_n^2}\to 0, \text{ as } n\to\infty$$

see, Hall (1984).

Lemma A.4 *As $n\to\infty$ in (A.8) one has*

(A.9) $$A_n = \frac{2\sigma_T^2}{h^{4\nu+1}n^4}+O\left(\frac{1}{h^{4\nu}n^4}\right)$$

Proof. We start with the definition of A_n and equation (A.6) in Lemma A.2

$$A_n = E\left[\frac{1}{h^{2\nu+1}n^2}(K_\nu^* * K_\nu^*)\left(\frac{X_{1\alpha}-X_{2\alpha}}{h}\right)\frac{\varphi(\overline{X}_1)\varphi(\overline{X}_2)\sigma(X_1)\sigma(X_2)}{\varphi(X_{1\alpha},\overline{X}_1)\varphi(X_{1\alpha},\overline{X}_2)}\pi(X_{1\alpha})\{1+O_p(h)\}\right]^2$$

or

$$\frac{1}{h^{4\nu+2}n^4}\int\left[(K_\nu^* * K_\nu^*)\left(\frac{x_\alpha-y_\alpha}{h}\right)\frac{\varphi(\overline{x})\varphi(\overline{y})\sigma(x)\sigma(y)}{\varphi(x_\alpha,\overline{x})\varphi(x_\alpha,\overline{y})}\pi(x_\alpha)\{1+O_p(h)\}\right]^2$$
$$\varphi(x_\alpha,\overline{x})\varphi(y_\alpha,\overline{y})dx_\alpha d\overline{x}dy_\alpha d\overline{y}$$

and which equals, by change of variables $y_\alpha = x_\alpha + hu$

$$\frac{1}{h^{4\nu+1}n^4}\int\left[(K_\nu^* * K_\nu^*)(u)\frac{\varphi(\overline{x})\varphi(\overline{y})\sigma(x)\sigma(x_\alpha+hu,\overline{y})}{\varphi(x_\alpha,\overline{x})\varphi(x_\alpha,\overline{y})}\pi(x_\alpha)\right]^2$$
$$\varphi(x_\alpha,\overline{x})\varphi(x_\alpha+hu,\overline{y})dx_\alpha d\overline{x}dud\overline{y}\ \{1+O(h)\}$$
$$= \frac{\|(K_\nu^* * K_\nu^*)\|_{L^2}^2}{h^{4\nu+1}n^4}\int\frac{\varphi(\overline{x})^2\varphi^2(\overline{y})\sigma^2(x)\sigma^2(x_\alpha,\overline{y})}{\varphi(x_\alpha,\overline{x})\varphi(x_\alpha,\overline{y})}\pi^2(x_\alpha)dx_\alpha d\overline{x}d\overline{y}\ \{1+O(h)\}$$
$$= \frac{2}{h^{4\nu+1}n^4}\frac{\left\|K_\nu^{*(2)}\right\|_{L^2}^2}{2}\int\left\{\int\frac{\sigma^2(x_\alpha,\overline{x})\varphi^2(\overline{x})}{\varphi(x)}d\overline{x}\right\}^2\pi(x_\alpha)^2dx_\alpha\{1+O(h)\}$$
$$= \frac{2\sigma_T^2}{h^{4\nu+1}n^4}+O\left(\frac{1}{h^{4\nu}n^4}\right)$$

Q.E.D.

Lemma A.5 *As $n\to\infty$, in (A.8) one has*

$$\text{(A.10)}\qquad B_n = \frac{\left\|K_\nu^{*(2)}\right\|_{L^4}^4}{h^{8\nu+3}n^8}\int\left\{\int\frac{\sigma^4(x)\varphi^4(\overline{x})}{\varphi^3(x)}d\overline{x}\right\}^2\pi^4(x_\alpha)dx_\alpha+O\left(\frac{1}{h^{8\nu+2}n^8}\right)$$

Proof. Like for A_n, we start with the definition of B_n and equation (A.6) in Lemma A.2

$$B_n = E\left[\frac{1}{h^{2\nu+1}n^2}(K_\nu^* * K_\nu^*)\left(\frac{X_{1\alpha}-X_{2\alpha}}{h}\right)\frac{\varphi(\overline{X}_1)\varphi(\overline{X}_2)\sigma(X_1)\sigma(X_2)}{\varphi(X_{1\alpha},\overline{X}_1)\varphi(X_{1\alpha},\overline{X}_2)}\pi(X_{1\alpha})\{1+O_p(h)\}\right]^4$$

or

$$\frac{1}{h^{8\nu+4}n^8}\int\left[(K_\nu^* * K_\nu^*)\left(\frac{x_\alpha-y_\alpha}{h}\right)\frac{\varphi(\overline{x})\varphi(\overline{y})\sigma(x)\sigma(y)}{\varphi(x_\alpha,\overline{x})\varphi(x_\alpha,\overline{y})}\pi(x_\alpha)\{1+O_p(h)\}\right]^4\varphi(x_\alpha,\overline{x})\varphi(y_\alpha,\overline{y})dx_\alpha d\overline{x}dy_\alpha d\overline{y}$$

and which equals, by change of variables $y_\alpha = x_\alpha + hu$

$$\frac{1}{h^{8\nu+3}n^8}\int\left[(K_\nu^* * K_\nu^*)(u)\frac{\varphi(\overline{x})\varphi(\overline{y})\sigma(x)\sigma(x_\alpha+hu,\overline{y})}{\varphi(x_\alpha,\overline{x})\varphi(x_\alpha,\overline{y})}\pi(x_\alpha)\right]^4\varphi(x_\alpha,\overline{x})\varphi(x_\alpha+hu,\overline{y})dx_\alpha d\overline{x}dud\overline{y}\ \{1+O(h)\}$$
$$= \frac{\|(K_\nu^* * K_\nu^*)\|_{L^4}^4}{h^{8\nu+3}n^8}\int\frac{\varphi(\overline{x})^4\varphi^4(\overline{y})\sigma^4(x)\sigma^4(x_\alpha,\overline{y})}{\varphi^3(x_\alpha,\overline{x})\varphi^3(x_\alpha,\overline{y})}\pi^4(x_\alpha)dx_\alpha d\overline{x}d\overline{y}\ \{1+O(h)\}$$
$$= \frac{\left\|K_\nu^{*(2)}\right\|_{L^4}^4}{h^{8\nu+3}n^8}\int\left\{\int\frac{\sigma^4(x_\alpha,\overline{x})\varphi^4(\overline{x})}{\varphi^3(x)}d\overline{x}\right\}^2\pi^4(x_\alpha)dx_\alpha\{1+O(h)\}$$

Q.E.D.

Now we want to calculate $C_n = E\left[J_n(\epsilon_1,X_1,\epsilon_2,X_2)\right]^2$, where

$$J_n(\epsilon,X,\delta,Y) = E_1\left[\epsilon_1\epsilon A(X_1,X)\sigma(X_1)\sigma(X)\epsilon_1\delta A(X_1,Y)\sigma(X_1)\sigma(Y)\right].$$

Lemma A.6 *It holds that*

$$J_n(\epsilon, X, \delta, Y) = \frac{\epsilon\delta\overline{\varphi}(\overline{X})\sigma(X)\overline{\varphi}(\overline{Y})\sigma(Y)\pi^2(X_\alpha)}{h^{4\nu+1}n^4\varphi(X_\alpha, \overline{X})\varphi(X_\alpha, \overline{Y})} K_\nu^{*(4)}\left(\frac{Y_\alpha - X_\alpha}{h}\right)\int \frac{\overline{\varphi}^2(\overline{x})\sigma^2(X_\alpha, \overline{x})}{\varphi(X_\alpha, \overline{x})} d\overline{x}\left\{1 + O_p(h)\right\}$$

(A.11)

Proof. By definition of J_n and equation (A.6) in Lemma A.2

$$J_n(\epsilon, X, \delta, Y) = \frac{\epsilon\delta}{h^{4\nu+2}n^4} E\left[K_\nu^{*(2)}\left(\frac{X_{1\alpha} - X_\alpha}{h}\right)\frac{\overline{\varphi}(\overline{X}_1)\overline{\varphi}(\overline{X})\sigma(X_1)\sigma(X)}{\varphi(X_{1\alpha}, \overline{X}_1)\varphi(X_{1\alpha}, \overline{X})}\pi(X_{1\alpha})\right.$$

$$\left. K_\nu^{*(2)}\left(\frac{X_{1\alpha} - Y_\alpha}{h}\right)\frac{\overline{\varphi}(\overline{X}_1)\overline{\varphi}(\overline{Y})\sigma(X_1)\sigma(Y)}{\varphi(X_{1\alpha}, \overline{X}_1)\varphi(X_{1\alpha}, \overline{Y})}\pi(X_{1\alpha})\left\{1 + O_p(h)\right\}\right]$$

or

$$J_n(\epsilon, X, \delta, Y) = \frac{\epsilon\delta}{h^{4\nu+2}n^4}\int K_\nu^{*(2)}\left(\frac{x_\alpha - X_\alpha}{h}\right)\frac{\overline{\varphi}(\overline{x})\overline{\varphi}(\overline{X})\sigma(X)}{\varphi(x_\alpha, \overline{x})\varphi(x_\alpha, \overline{X})}\sigma^2(x)$$

$$K_\nu^{*(2)}\left(\frac{x_\alpha - Y_\alpha}{h}\right)\frac{\overline{\varphi}(\overline{x})\overline{\varphi}(\overline{Y})\sigma(Y)}{\varphi(x_\alpha, \overline{x})\varphi(x_\alpha, \overline{Y})}\pi^2(x_\alpha)\varphi(x_\alpha, \overline{x})dx_\alpha d\overline{x}\left\{1 + O_p(h)\right\}$$

which, by a change of variable $x_\alpha = X_\alpha + hu$, becomes

$$J_n(\epsilon, X, \delta, Y) = \frac{\epsilon\delta}{h^{4\nu+1}n^4}\int K_\nu^{*(2)}(u)\frac{\overline{\varphi}(\overline{x})\overline{\varphi}(\overline{X})\sigma(X)}{\varphi(X_\alpha + hu, \overline{x})\varphi(X_\alpha + hu, \overline{X})}\sigma^2(X_\alpha + hu, \overline{x})$$

with

$$K_\nu^{*(2)}\left(\frac{X_\alpha - Y_\alpha}{h} + u\right)\frac{\overline{\varphi}(\overline{x})\overline{\varphi}(\overline{Y})\sigma(Y)}{\varphi(X_\alpha + hu, \overline{x})\varphi(X_\alpha + hu, \overline{Y})}\pi^2(X_\alpha + hu)\varphi(X_\alpha + hu, \overline{x})dud\overline{x}\left\{1 + O_p(h)\right\}$$

and thus $J_n(\epsilon, X, \delta, Y) =$

$$\frac{\epsilon\delta\overline{\varphi}(\overline{X})\sigma(X)\overline{\varphi}(\overline{Y})\sigma(Y)\pi^2(X_\alpha)}{h^{4\nu+1}n^4\varphi(X_\alpha, \overline{X})\varphi(X_\alpha, \overline{Y})}(K_\nu^{*(4)})\left(\frac{Y_\alpha - X_\alpha}{h}\right)\int \frac{\overline{\varphi}^2(\overline{x})\sigma^2(X_\alpha, \overline{x})}{\varphi(X_\alpha, \overline{x})}d\overline{x}\left\{1 + O_p(h)\right\}$$

Q.E.D.

Lemma A.7

$$\text{(A.12)} \qquad C_n = \frac{\left\|K_\nu^{*(4)}\right\|_{L^2}^2}{h^{8\nu+1}n^8}\int\left\{\int \frac{\overline{\varphi}(\overline{x})^2\sigma(x_\alpha, \overline{x})^2}{\varphi(x_\alpha, \overline{x})}d\overline{x}\right\}^4 \pi^4(x_\alpha)dx_\alpha + O\left(\frac{1}{h^{8\nu}n^8}\right)$$

Proof. By definition, the number C_n is

$$E\left[J_n(\epsilon_1, X_1, \epsilon_2, X_2)\right]^2 =$$

$$E\left[\epsilon_1\epsilon_2\overline{\varphi}(\overline{X}_1)\sigma(X_1)\overline{\varphi}(\overline{X}_2)\sigma(X_2)\pi^2(X_{1\alpha})K_\nu^{*(4)}\left(\frac{X_{1\alpha} - X_{2\alpha}}{h}\right)\int \frac{\overline{\varphi}(\overline{x})^2\sigma(X_{1\alpha}, \overline{x})^2}{\varphi(X_{1\alpha}, \overline{x})}d\overline{x}\right.$$

$$\left. h^{-(4\nu+1)}n^{-4}\varphi^{-1}(X_{1\alpha}, \overline{X}_1)\varphi^{-1}(X_{1\alpha}, \overline{X}_2)\left\{1 + O_p(h)\right\}\right]^2$$

or

$$\int \overline{\varphi}^2(\overline{x}_1)\sigma^2(x_1)\overline{\varphi}^2(\overline{x}_2)\sigma^2(x_2)\pi^4(x_{1\alpha})K_\nu^{*(4)}\left(\frac{x_{1\alpha} - x_{2\alpha}}{h}\right)^2\left\{\int \frac{\overline{\varphi}(\overline{x})^2\sigma(x_{1\alpha}, \overline{x})^2}{\varphi(x_{1\alpha}, \overline{x})}d\overline{x}\right\}^2$$

$$h^{-2(4\nu+1)}n^{-8}\varphi^{-2}(x_{1\alpha},\overline{x}_1)\varphi^{-2}(x_{1\alpha},\overline{x}_2)\varphi(x_{1\alpha},\overline{x}_1)\varphi(x_{2\alpha},\overline{x}_2)dx_{1\alpha}d\overline{x}_1dx_{2\alpha}d\overline{x}_2\left\{1+O(h)\right\}$$

and by changing the variable $x_{1\alpha}-x_{2\alpha}=hu$, the above equals

$$\int\overline{\varphi}^2(\overline{x}_1)\sigma^2(x_1)\overline{\varphi}^2(\overline{x}_2)\sigma^2(x_{1\alpha}-hu,\overline{x}_2)\pi^4(x_{1\alpha})K_\nu^{*(4)}(u)^2\left\{\int\frac{\overline{\varphi}(\overline{x})^2\sigma(x_{1\alpha},\overline{x})^2}{\varphi(x_{1\alpha},\overline{x})}d\overline{x}\right\}^2$$

$$h^{-2(4\nu+1)+1}n^{-8}\varphi^{-2}(x_{1\alpha},\overline{x}_1)\varphi^{-2}(x_{1\alpha},\overline{x}_2)\varphi(x_{1\alpha},\overline{x}_1)\varphi(x_{1\alpha}-hu,\overline{x}_2)dx_{1\alpha}d\overline{x}_1dud\overline{x}_2\left\{1+O(h)\right\}$$

$$=\frac{\left\|K_\nu^{*(4)}\right\|_{L^2}^2}{h^{8\nu+1}n^8}\int\frac{\overline{\varphi}^2(\overline{x}_1)\sigma^2(x_1)\overline{\varphi}^2(\overline{x}_2)\sigma^2(x_{1\alpha},\overline{x}_2)\pi^4(x_{1\alpha})}{\varphi(x_{1\alpha},\overline{x}_1)\varphi(x_{1\alpha},\overline{x}_2)}$$

$$\left\{\int\frac{\overline{\varphi}(\overline{x})^2\sigma(x_{1\alpha},\overline{x})^2}{\varphi(x_{1\alpha},\overline{x})}d\overline{x}\right\}^2dx_{1\alpha}d\overline{x}_1d\overline{x}_2\left\{1+O(h)\right\}$$

$$=\frac{\left\|K_\nu^{*(4)}\right\|_{L^2}^2}{h^{8\nu+1}n^8}\int\left\{\int\frac{\overline{\varphi}(\overline{x})^2\sigma(x_{1\alpha},\overline{x})^2}{\varphi(x_{1\alpha},\overline{x})}d\overline{x}\right\}^4\pi^4(x_{1\alpha})dx_{1\alpha}\left\{1+O(h)\right\}$$

$$=\frac{\left\|K_\nu^{*(4)}\right\|_{L^2}^2}{h^{8\nu+1}n^8}\int\left\{\int\frac{\overline{\varphi}(\overline{x})^2\sigma(x_\alpha,\overline{x})^2}{\varphi(x_\alpha,\overline{x})}d\overline{x}\right\}^4\pi^4(x_\alpha)dx_\alpha+O\left(\frac{1}{h^{8\nu}n^8}\right)$$

Q.E.D.

Lemma A.8 *As $n\to\infty$ it holds*

$$\sqrt{h^{4\nu+1}n^2}Q_2\xrightarrow{D}N\left(0,\sigma_T^2\right)\tag{A.13}$$

Proof. we have established in (A.9), (A.10), and (A.12) that $A_n\propto\frac{1}{h^{4\nu+1}n^4}$, $B_n\propto\frac{1}{h^{8\nu+3}n^8}$, and $C_n\propto\frac{1}{h^{8\nu+1}n^8}$, and hence

$$\frac{C_n+\frac{1}{n}B_n}{A_n^2}=O(h+\frac{1}{nh})\to 0\text{ , as }n\to\infty\text{ .}$$

Therefore by the central limit theorem for non-degenerate U-statistic as in Hall (1984), $\sqrt{h^{4\nu+1}n^2}Q_2$ is asymptotically normal with asymptotic variance

$$\frac{n^2}{2}h^{4\nu+1}n^2A_n=\frac{n^2}{2}h^{4\nu+1}n^2\frac{2\sigma_T^2}{h^{4\nu+1}n^4}=\sigma_T^2$$

Q.E.D.

Proof of Theorem 2. Now combining the results on Q_1 and Q_2, namely (A.7) in Lemma A.3 and (A.13) in Lemma A.8, plus equation (A.2), we obtain the equation (16) in Theorem 2. Q.E.D.

References

ANDREWS, D.W.K. AND WHANG, Y.-J. (1990) Additive interactive regression models: circumvention of the curse of dimensionality. *Econometric Theory,* **6** 466-479.

FAN, J., HÄRDLE, W. AND MAMMEN, E. (1998) Direct estimation of low dimensional components in additive models. *Annals of Statistics,* **26** 943-971.

GOZALO, P.L. AND LINTON, O.B. (1999) Testing additivity in generalized nonparametric regression models. *Preprint, Yale University, USA*

HALL, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators, *Journal of Multivariate Analysis,* **14** 1-16.

HÄRDLE, H., S. HUET, E. MAMMEN, AND S. SPERLICH (2000) Semiparametric additive indices for binary response and generalized additive models. *Working Paper, Carlos III de Madrid, Spain.*

HÄRDLE, W. AND MAMMEN, E. (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics,* **21** 1926-1947.

HÄRDLE, W. AND MARRON, J.S. (1991) Bootstrap Simultaneous Error Bars for Nonparametric Regression. *Annals of Statistics,* **19** 778-796.

HASTIE, T. J. AND TIBSHIRANI, R. J. (1990) *Generalized Additive Models.* Chapman and Hall: London

HJELLVIK, V., YAO, Q. AND TJØSTHEIM, D. (1998) Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference,* **68** 295-321.

LINTON, O. B. AND HÄRDLE, W. (1996) Estimation of additive regression models with known links. *Biometrika,* **83** 529-540.

LINTON, O.B. AND NIELSEN, J.P. (1995) A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika,* **82** 93-101.

LIU, R. (1988). Bootstrap procedures under some non i.i.d. models. *Annals of Statistics* **16** 1696-1708.

MAMMEN, E., LINTON, O.B. AND NIELSEN, J.P. (1999) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics,* **27** 1443-1490.

NIELSEN, J.P. AND LINTON, O.B. (1997) An optimization interpretation of integration and backfitting estimators for separable nonparametric models. *Journal of the Royal Statistical Society, Series B,* **60** 217-222.

RUPPERT, D. AND WAND, M. P. (1994) Multivariate locally weighted least squares regression. *Annals of Statistics,* **22** 1346-1370.

SEVERANCE-LOSSIN, E. AND SPERLICH, S. (1999) Estimation of derivatives for additive separable models. *Statistics,* **33** 241-265.

SPERLICH, S., LINTON, O.B. AND HÄRDLE, W. (1999) Integration and Backfitting methods in additive models: Finite sample properties and comparison. *Test,* **8** 419-458.

SPERLICH, S., TJØSTHEIM, D. AND YANG , L. (2000) Nonparametric Estimation and Testing of Interaction in Additive Models. revision of *Working Paper 99-85 (34), Carlos III de Madrid, Spain, forthcoming in Econometric Theory.*

STONE, C.J. (1985) Additive regression and other nonparametric models. *Annals of Statistics,* **13** 689-705.

STONE, C.J. (1986) The dimensionality reduction principle for generalized additive models. *Annals of Statistics,* **14** 90-606.

TJØSTHEIM, D. AND AUESTAD, B.H. (1994) Nonparametric identification of nonlinear time series: projections. *Journal of American Statistical Association*, **89** 1398-1409.

WU, C. F. J. (1986) Jacknife, bootstrap and other resampling methods in regression analysis (with discussion). *Annals of Statistics*, **14** 1261-1350.